ylenetetrahydrofolate reductase and neural tube defects. Lancet 348:58
Rees DC, Cox MJ, Clegg JB (1995) World distribution of factor V Leiden. Lancet 346:1133–1134
Sever LE (1982) An epidemiologic study of neural tube defects in Los Angeles County. II. Etiologic factors in an area with low prevalence at birth. Teratology 25:323–334
Smithells RW, Sheppard S, Schorah CJ, Seller MJ, Nevin NC, Harris R, Read AP et al (1980) Possible prevention of neural-tube defects by periconceptional vitamin supplementation. Lancet 1:339–40
Sohda S, Arinami T, Hamada H, Yamada N, Hamaguchi H, Kubo T (1997) Methylenetetrahydrofolate reductase polymorphism and pre-eclampsia. J Med Genet 34:525–526
Stevenson RE, Schwartz CE, Du Y-Z, Adams MJ Jr. (1997) Differences in methylenetetrahydrofolate reductase genotype frequencies between whites and blacks. Am J Hum Genet 60:229–230
Thompson EA, Neel JV (1997) Allelic disequilibrium and allele frequency distribution as a function of social and demographic history. Am J Hum Genet 60:197–204
van der Put NMJ, Eskes TKAB, Blom HJ (1997) Is the common 677C→T mutation in the methylenetetrahydrofolate reductase gene a risk factor for neural tube defects? A meta-analysis. Q J Med 90:111–115
Wilcken DEL (1997) MTHFR 677C-T mutation, folate intake, neural-tube defect and risk of cardiovascular disease. Lancet 350:603–604

Address for correspondence and reprints: Ms. Julie A. Schneider, MRC Molecular Haematology Unit, Institute of Molecular Medicine, The John Radcliffe, Headley Way, Headington, Oxford OX3 9DS, United Kingdom. E-mail: julie.schneider@bioanth.ox.ac.uk

0002-9297/98/6205-0033$02.00

Am. J. Hum. Genet. 62:1260–1262, 1998

Mutational Mechanisms for Generating Microsatellite Allele-Frequency Distributions: An Analysis of 4,558 Markers

To the Editor:

Genomewide linkage searches have been facilitated by the development of panels of microsatellite markers that are widely distributed throughout the genome and are highly polymorphic. Population geneticists have investigated mutational mechanisms for generating new microsatellite alleles, considered infinite allele models (Shriver et al. 1993), single-step mutation models (Shriver et al. 1993; Valdes et al. 1993; Di Rienzo et al. 1994; Deka et al. 1995), and multistep mutation models (Di Rienzo et al. 1994; Kimmel and Chakraborty 1996; Chakraborty et al. 1997), and compared theoretical predictions of various parameters such as number of alleles and degree of heterozygosity with real-world observations. These studies have, in general, examined relatively few markers with data derived from several populations. Here we report an analysis of the extensive, publicly available data for the Généthon $(AC)_n$ microsatellite markers (Dib et al. 1996) to investigate (1) the distribution of allele frequencies and (2) the distribution of size differences between alleles, for an idealized microsatellite marker, and to examine the implications of these observations for the underlying mutational model.

The Fondation Jean Dausset–CEPH database (version 8.1) was downloaded from the FTP server (ftp.cephb.fr). Five thousand sixty-three autosomal Généthon $(AC)_n$ microsatellite markers were identified by the nominal prefix "AFM;" 329 markers were eliminated from subsequent analysis, since the difference in size in base pairs between alleles was not an exact multiple of two. The genotypes of 22 unrelated founders of families 1332, 1347, 1362, 1413, and 1416, all of which originate in Utah, were then compiled for subsequent analysis. Markers were grouped by the number of different-sized alleles found in the sample of 44 Utah chromosomes; alleles were then ranked by size (bp) and the mean frequencies of the ranked alleles for 4,558 markers with between 3 and 11 alleles are plotted in figure 1*A*. The frequency distribution traces a distinctive, asymmetrical pattern that follows a function of the number of alleles. As expected, the mode allele lies midway in rank, and its frequency decreases with the total number of alleles. More noteworthy, we observe that the frequency distributions are all positively skewed (coefficients of skewness range from 0.074 to 0.211), the data are significantly different ($P < .01$) from a random sample drawn from a normal distribution using the Kolmogorov test (performed using the SAS UNIVARIATE procedure [SAS Institute 1990]). The frequency distributions for an additional 176 markers with between 12 and 22 different alleles are not shown, because there were insufficient numbers within each size class, leading to excessive variability.

Computer simulation studies were performed to explore plausible mutational mechanisms that underlie the asymmetrical distribution of mean frequencies of the ranked alleles. Models were based on the Fisher-Wright genetic drift model, in which $2N$ chromosomes were sampled with replacement from a diploid population of size N. Mutations at a rate ν were assigned that replace an allele of size S (measured as number of dinucleotides) with a larger or smaller allele. Markers were assumed to be unlinked and in linkage equilibrium. We examined a single-step mutation model (SSMM) in which newly mutated alleles have size $S + 1$ or $S - 1$ with equal probability (i.e., one dinucleotide repeat motif larger or smaller). We also examined multiple-step mutation models (MSMM) in which new alleles have size $S + n$ or

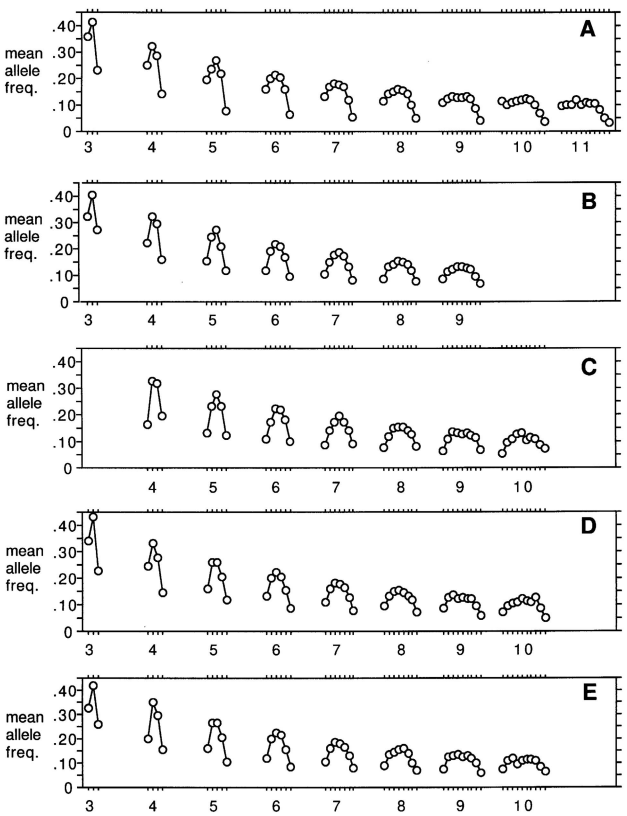

Figure 1 Distributions of mean frequencies for ranked alleles for microsatellite markers that are grouped by the number of alleles (indicated on the abscissa). *A*, Fesults from an analysis of 4,558 Généthon $(AC)_n$ microsatellite markers. *B*, Results for a simulation study for a multiple-step mutation model (MSMM) in which the initial number of repeat units (start size *S*) is five (no mutational bias). *C*, Results for a similar simulation study in which the start size *S* is 50 (no mutational bias). *D*, Results for a MSMM simulation study with a 2:1 mutational bias in favor of larger alleles in which the start size *S* is 5; *E*, Results for a similar simulation study in which the start size *S* is 50.

$S - n$ (where n is the number of dinucleotide repeat motifs) with alternatively equal probabilities (.5:.5) or with unequal probabilities (.666:.333) to introduce a mutational bias in favor of larger alleles. The relative mutation rates for step size n were modeled by a standard geometric distribution in which the relative mutation rate (ν_n) is defined as $\nu_n = m/(m+1)^n$, which has the useful property, when $m > 0$, that

$$\sum_{n=1}^{\infty} \frac{m}{(m+1)^n} = 1 \ .$$

These models define a natural lower "absorbing boundary" for the number of dinucleotide repeats contained within a microsatellite allele (i.e., mutations to zero or negative sizes are not permitted, so $n \geqslant 1$). Each simulation exercise commenced with a single allele of size *S*, where *S* ranged from 1 to 50. As the Généthon markers were selected to ensure that they were highly polymorphic, replicates were discarded if the heterozygosity for the marker at the end of the simulations did not exceed 50%. Simulations were carried out for a range of values of $4N\nu$, number of generations, and allele start size (*S*) for both the SSMM and the MSMM (and *m* for the MSMM simulations).

Figure 1 shows the mean frequencies of the ranked alleles in a random sample of 44 chromosomes following a simulation study of 5,000 markers over 20,000 generations under the MSMM with no mutational bias, with $4N\nu = 4$ and $m = 2$ and with the start size $S = 5$ (fig. 1*B*) or $S = 50$ (fig. 1*C*). It is evident by inspection that the simulated distribution with a start size $S = 5$ is positively skewed (fig. 1*B*), but when $S = 50$ there is virtually no skewness (fig. 1*C*). This demonstrates that asymmetry can result from a simple absorbing boundary model provided that the start size *S* is close to the boundary; other simulations with a range of start sizes indicate that the degree of skewing is roughly inversely proportional to *S*. Figures 1*D* and 1*E* show the results from a MSMM with a 2:1 mutational bias in favor of larger alleles with the start size $S = 5$ (fig. 1*D*) or with the start size $S = 50$ (fig. 1*E*). These latter results show that a mutational bias model (fig. 1*E*) can result in skewed allele frequency distributions. However, we note that even the simulation with the start size $S = 5$ and mutational bias (fig. 1*D*) does not appear to be as skewed as the Généthon data (fig. 1*A*).

The parameter $4N\nu$ was selected to model a randomly mating population of size 2,000 and a mutation rate $\nu = .0005$. This latter value is consistent with microsatellite mutation rate estimates reported in the literature which range from 10^{-2} to 10^{-5}. The simulated distributions of the number of alleles and heterozygosity were broadly comparable to those observed in the Généthon data (further details and results from additional simulations are available on the homepage http://well.ox.ac.uk/~mfarrall/microsatellite.html).

We have also examined the distribution of size differences of alleles, expressed as the number of dinucleotide repeat units, between adjacently ranked alleles, and we list the results from our analysis of the Généthon data and the SSMM and MSMM simulation studies for 20,000 generations in table 1. We note that, by inspection, the distribution of the size differences between adjacent ranked alleles in the Généthon sample is very closely matched by the MSMM (with no mutational bias) simulated data (where $m = 2$), while the SSMM underestimates the frequency of large differences in size between adjacently ranked alleles.

Table 1

Frequency Distribution of Size Differences between Microsatellite Alleles

No. of Repeats	Généthon	SSMM	MSMM
1	.812	.936	.810
2	.111	.047	.126
3	.036	.011	.037
4	.016	.004	.014
5	.011	.001	.007
6	.005	.001	.003
7	.003	.000	.002
8	.002	.000	.001
9	.001	.000	.001
10	.001	.000	.000

NOTE.—"No. of Repeats" indicates the difference in the number of dinucleotide repeats between adjacent alleles that have been ranked by their absolute size. SSMM = single-step mutation model. MSMM = multiple-step mutation model ($m = 2$), no mutational bias.

We have found that microsatellite allele frequency distributions tend to be positively skewed in favor of longer alleles, which is in agreement with an analysis of CAG repeats in the Huntington disease gene (Rubinsztein et al. 1994) and the unpublished results of W. Amos and D. Rubinsztein (cited in Rubinsztein et al. 1995). Our computer simulation results suggest that the underlying mutational model for generating new microsatellite alleles is likely to be asymmetrical and multistep. MSMM models with an absorbing boundary or with a mutational bias in favor of larger alleles can generate allele distributions that closely resemble those observed in the Généthon data. Models of directional evolution that result from mutational bias have been recently discussed (Rubinsztein et al. 1995; Primmer et al. 1996). Finally, the empirical mean frequencies of the ranked alleles derived from the Généthon analysis provide useful prior distributions to those applying Bayesian smoothing techniques (Lange 1997) to $(AC)_n$ microsatellite allele frequency estimates.

MARTIN FARRALL[1] AND DANIEL E. WEEKS[1,2]
[1]*Statistical Genetics Group, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford; and* [2]*Department of Human Genetics, University of Pittsburgh, Pittsburgh*

References

Chakraborty R, Kimmel M, Stivers DN, Davison LJ, Deka R (1997) Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. Proc Natl Acad Sci USA 94: 1041–1046

Deka R, Jin L, Shriver MD, Yu LM, DeCroo S, Hundrieser J, Bunker CH, et al (1995) Population genetics of dinucleotide $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ polymorphisms in world populations. Am J Hum Genet 56:461–474

Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, et al (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380:152–154

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci USA 91:3166–3170

Kimmel M, Chakraborty R (1996) Measures of variation at DNA repeat loci under a general stepwise mutation model. Theor Popul Biol 50:345–367

Lange K (1997) Mathematical and statistical methods for genetic analysis. Springer, Berlin

Primmer CR, Ellegren H, Saino N, Møller AP (1996) Directional evolution in germline microsatellite mutations. Nat Genet 13:391–393

Rubinsztein DC, Amos W, Leggo J, Goodburn S, Jain S, Li S-H, Margolis RL, et al (1995) Microsatellite evolution—evidence for directionality and variation in rate between species. Nat Genet 10:337–343

Rubinsztein DC, Amos W, Leggo J, Goodburn S, Ramesar S, Old J, Bontrop R, et al (1994) Mutational bias provides a model for the evolution of Huntington's disease and predicts a general increase in disease prevalence. Nat Genet 7: 525–530

SAS Institute (1990) SAS procedures guide, version 6. 3d ed. SAS Institute, Cary NC

Shriver MD, Jin L, Chakraborty R, Boerwinkle E (1993) VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. Genetics 134:983–993

Valdes AM, Slatkin M, Freimer NB (1993) Allele frequencies at microsatellite loci: the stepwise mutation model revisited. Genetics 133:737–749

Address for correspondence and reprints: Dr. Martin Farrall, Statistical Genetics group, Wellcome Trust Centre for Human Genetics, University of Oxford, Windmill Road, Oxford, OX3 7BN, United Kingdom. E-mail: mfarrall@well.ox.ac.uk

0002-9297/98/6205-0034$02.00

Am. J. Hum. Genet. 62:1262–1265, 1998

Some Underlooked Properties of the Multifactorial/ Threshold Model

To the Editor:

Some years ago, when the multifactorial/threshold (MFT) model was beginning to be recognized (by some) as a useful way of thinking about the causes of common congenital malformations, I noted one of its implications. "It follows from the MFT model that in conditions appearing more often in one sex than the other, the sex ratio should change as the frequency changes" (Fraser 1971, p. 90). I suggested that such changes in the sex